Artificial Software Diversification for WebAssembly

JAVIER CABRERA-ARTEAGA

Doctoral Thesis
Supervised by
Benoit Baudry and Martin Monperrus
Stockholm, Sweden, 2023

KTH Royal Institute of Technology
School of Electrical Engineering and Computer Science
Division of Software and Computer Systems
SE-10044 Stockholm
Sweden

TRITA-EECS-AVL-2020:4
ISBN 100-

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av Teknologie doktorexamen i elektroteknik i .

Tryck: Universitetsservice US AB

Abstract

[1]

Keywords: Lorem, Ipsum, Dolor, Sit, Amet

Sammanfattning

[1]

LIST OF PAPERS

1. ***Superoptimization of WebAssembly Bytecode***
   Javier Cabrera-Arteaga,Shrinish Donde, Jian Gu, Orestis Floros, Lucas Satabin, Benoit Baudry, Martin Monperrus
   Conference Companion of the 4th International Conference on Art, Science, and Engineering of Programming (Programming 2021), MoreVMs
   `https://doi.org/10.1145/3397537.3397567`

2. ***CROW: Code Diversification for WebAssembly***
   Javier Cabrera-Arteaga, Orestis Floros, Oscar Vera-Pérez, Benoit Baudry, Martin Monperrus
   Network and Distributed System Security Symposium (NDSS 2021), MADWeb
   `https://doi.org/10.14722/madweb.2021.23004`

3. ***Multi-Variant Execution at the Edge***
   Javier Cabrera-Arteaga,Pierre Laperdrix, Martin Monperrus, Benoit Baudry
   Conference on Computer and Communications Security (CCS 2022), Moving Target Defense (MTD)
   `https://dl.acm.org/doi/abs/10.1145/3560828.3564007`

4. ***WebAssembly Diversification for Malware Evasion***
   Javier Cabrera-Arteaga,Tim Toady, Martin Monperrus, Benoit Baudry
   Computers & Security, Volume 131, 2023
   `https://www.sciencedirect.com/science/article/pii/S0167404823002067`

5. ***Wasm-mutate: Fast and Effective Binary Diversification for WebAssembly***
   Javier Cabrera-Arteaga, Nick Fitzgerald, Martin Monperrus, Benoit Baudry

6. ***Scalable Comparison of JavaScript V8 Bytecode Traces***
   Javier Cabrera-Arteaga,Martin Monperrus, Benoit Baudry
   11th ACM SIGPLAN International Workshop on Virtual Machines and Intermediate Languages (SPLASH 2019)
   `https://doi.org/10.1145/3358504.3361228`

ACKNOWLEDGEMENT

[1]

List of commonly used acronyms:

AE Acronym examples

Contents

01 INTRODUCTION

1.1 Background

TODO Motivate with the open challenges.

1.2 Problem statement

TODO Problem statement **TODO** Set the requirements as R1, R2, then map each contribution to them.

1.3 Automatic Software diversification requirements

1. **1:** **TODO** Requirement 1

1.4 List of contributions

C1: Methodology contribution: We propose a methodology for generating software diversification for WebAssembly and the assessment of the generated diversity.

C2: Theoretical contribution: We propose theoretical foundation in order to improve Software Diversification for WebAssembly.

C3: Automatic diversity generation for WebAssembly: We generate WebAssembly program variants.

C4: Software Diversity for Defensive Purposes: We assess how generated WebAssembly program variants could be used for defensive purposes.

C5: Software Diversity for Offensives Purposes: We assess how generated WebAssembly program variants could be used for offensive purposes, yet improving security systems.

C6: Software Artifacts: We provide software artifacts for the research community to reproduce our results.

TODO Make multi column table

Contribution	Resarch papers				
	P1	P2	P3	P4	P5
C1	x	x		x	x
C2	x	x			
C3	x	x	x		
C4	x	x	x		
C5			x		
C6	x	x	x	x	x

Table 1.1: Mapping of the contributions to the research papers appended to this thesis.

■ 1.5 Summary of research papers

Paper 1: Superoptimization of WebAssembly Bytecode.

Paper 2: CROW: Code randomization for WebAssembly bytecode.

Paper 3: Multivariant execution at the Edge.

Paper 4: Wasm-mutate: Fast and efficient software diversification for WebAssembly.

Paper 5: WebAssembly Diversification for Malware evasion.

■ 1.6 Thesis outline

02 BACKGROUND AND STATE OF THE ART

2.1 WebAssembly

Roadmap

TODO Talk about the current roadmap. Components for WebAssembly. Polyglot inter process

WebAssembly toolchains

TODO Mention, stress the landscape of tools that involve Wasm. Include analysis tools, fuzzers, optimizers and malware detectors.

TODO End up motivating the need of Software Diversification for: testing and reliability.

2.2 Software diversification

2.3 Generating Software Diversification

Variants generation

Variants equivalence

Variants preservation

TODO We stress here our contributions with a new metric: variants preservation. We strees somehow that it is overlooked. Diversification, despite the stage at which is applied, for WebAssembly needs to go through a preservation study.

- 2.4 Exploiting Software Diversification
- 2.5 Defensive Diversification
- 2.6 Offensive Diversification
- 2.7 Contributions of this thesis to Software Diversification for WebAssembly

03

TECHNICAL CONTRIBUTIONS

3.1 Approach landscape

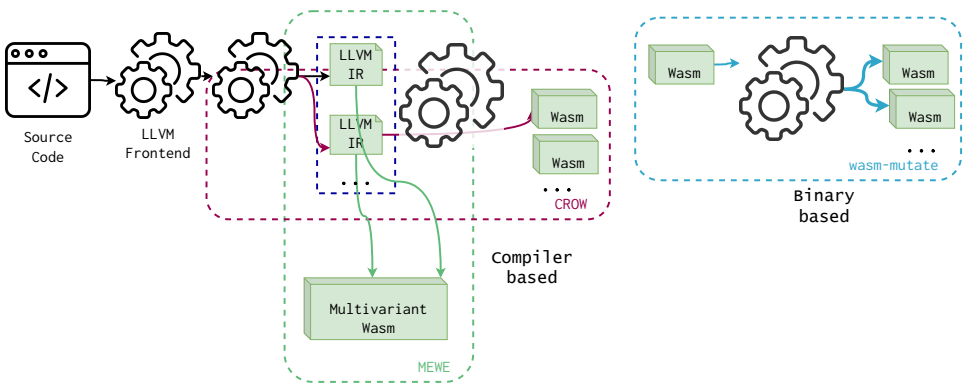

Figure 3.1: Approach landscape.

- 3.2 Compiler based approach
- CROW
- Constant inferring
- Disabling optimisations
- MEWE
- Multivariant binaries
- 3.3 Binary based approach
- wasm-mutate
- 3.4 Approaches comparison
- 3.5 Accompanying artifacts

04 EVALUATION

4.1 Research questions

RQ1: To what extent can we artificially generate program variants for WebAssembly?

RQ2: To what extent are the generated variants dynamically different?

RQ3: To what extent do the artificial variants exhibit different execution times on edge-cloud platforms?

RQ4: Defensive Diversification: ?

RQ5: Ofensive Diversification: ?

4.2 Experimental protocols

Metrics

New static metric. Diversification preservation.

4.3 Results

05 RESULTS AND DISCUSSION

- 5.1 Summary of technical contributions
- 5.2 Summary of empirical findings
- 5.3 Summary of empirical findings
- 5.4 Future Work

Part I

Included papers

SUPEROPTIMIZATION OF WEBASSEMBLY BYTECODE

Javier Cabrera-Arteaga, Shrinish Donde, Jian Gu, Orestis Floros, Lucas Satabin, Benoit Baudry, Martin Monperrus
Conference Companion of the 4th International Conference on Art, Science, and Engineering of Programming (Programming 2021), MoreVMs

https://doi.org/10.1145/3397537.3397567

CROW: CODE DIVERSIFICATION FOR WEBASSEMBLY

Javier Cabrera-Arteaga, Orestis Floros, Oscar Vera-Pérez, Benoit Baudry, Martin Monperrus
Network and Distributed System Security Symposium (NDSS 2021), MADWeb

https://doi.org/10.14722/madweb.2021.23004

MULTI-VARIANT EXECUTION AT THE EDGE

Javier Cabrera-Arteaga, Pierre Laperdrix, Martin Monperrus, Benoit Baudry
Conference on Computer and Communications Security (CCS 2022), Moving Target Defense (MTD)

https://dl.acm.org/doi/abs/10.1145/3560828.3564007

WEBASSEMBLY DIVERSIFICATION FOR MALWARE EVASION

Javier Cabrera-Arteaga, Tim Toady, Martin Monperrus, Benoit Baudry
Computers & Security, Volume 131, 2023

https://www.sciencedirect.com/science/article/pii/S0167404823002067

WASM-MUTATE: FAST AND EFFECTIVE BINARY DIVERSIFICATION FOR WEBASSEMBLY

Javier Cabrera-Arteaga, Nick Fitzgerald, Martin Monperrus, Benoit Baudry
Under revision

SCALABLE COMPARISON OF JAVASCRIPT V8 BYTECODE TRACES

Javier Cabrera-Arteaga, Martin Monperrus, Benoit Baudry
11th ACM SIGPLAN International Workshop on Virtual Machines and Intermediate Languages (SPLASH 2019)

https://doi.org/10.1145/3358504.3361228